

# 用于人形机器人的声音降噪方法

林剑峰, 冯 瑞

(复旦大学计算机科学与工程系肢体化智能实验室, 上海 200433)

**摘要:** 用于人形机器人的实时多通道降噪方法使机器人能够在实际应用中精确地进行后续的音频分类或语音识别。引入人形机器人具有类似人体的头相关传输函数的性质, 避免一般方法中在估计声传播信道模型时只考虑声达时延, 而忽略声散射和共振影响的局限性。该方法通过利用麦克风阵列阵元之间的声音传输信道的差异性, 削弱参考噪声和目标信号之间的相关性, 从而提升对消噪声后输出信号的信噪比。仿真表明, 对目标信号的信噪比提升达到 15 dB。

**关键词:** 人形机器人; 声音降噪; 麦克风阵列; 头相关传输函数

## Noise Reduction Method for Humanoid Robot

LIN Jian-feng, FENG Rui

(Embodied Intelligence Lab, Department of Computer Science and Engineering, Fudan University, Shanghai 200433)

**【Abstract】** A real time, multichannel, humanoid robot ready noise reduction method makes practical precise audio classification and speech recognition possible. Although scattering of sound plays an important role on humanoid robot, most conventional methods consider time delay as the only main factor for transferring model of sound propagation channel between microphones. By using characteristics of human like Head Related Transfer Function(HRTF) system, this method avoids the limitation of lackness of scattering effect. Simulations show that this method is able to improve desired sound's SNR by 15 dB.

**【Key words】** humanoid robot; noise reduction; microphone array; Head Related Transfer Function(HRTF)

### 1 概述

让机器人能够听懂人的口语和对特定声音做出反应是一项有意义的工作。然而, 现有语音识别或音频信号分类方法对于干扰噪声非常敏感, 在复杂环境中, 如会展、多人交谈等直接对音频信号进行识别难以获得令人满意的效果。因此, 如何对目标声音信号进行增强处理, 降低噪声影响, 提高目标声音信号信噪比, 是目前机器人领域亟待解决的问题。

目前, 利用麦克风阵列降噪要有自适应噪声对消和基于空间信息 2 类方法, 其中, 基于空间信息的方法又可分为 2 大类: 基于波束形成原理的方法和基于盲信号分离的方法。其中, 基于盲信号分离的各种算法复杂度较高, 不适合在机器人上实时运行。而波束形成方法又可以分为固定波束形成方法和自适应的波束形成方法, 其基本原理是利用阵列的空间信息对目标声源形成过滤波束来避免来自其他方向的噪声干扰。在用于机器人的降噪方法方面, 文献[1-2]取得了良好的效果, 但它们在利用空间信息时, 都采用了只有声达时延(Time Difference of Arrival, TDOA)为基础的声传播信道的模型, 而实际在机器人应用中, 散射和共振的效应也极大地影响了声传播信道的特性<sup>[3]</sup>, 因此, 可以加以改进。

自适应滤波对消降噪方法经过数十年的发展, 在很多领域获得成功应用<sup>[4]</sup>。其基本原理如图 1 所示。其中,  $x(n)$  是原始信号;  $a(n)$  是噪声;  $y(n)$  是受加性噪声干扰后的信号;  $\hat{b}(n)$  是加性噪声  $b(n)$  的估计;  $\hat{x}(n)$  是原始信号的估计。实际应用中,  $y(n)$  是采集的信号;  $\hat{x}(n)$  是增强后的目标信号。从中不难看出, 自适应滤波时, 至少要已知一路与噪声  $b(n)$  相关、与原始信号  $x(n)$  无关的参考噪声  $a(n)$ , 才能利用自适应滤波

获得最优的噪声估计  $\hat{b}(n)$ , 从而进行噪声消减。

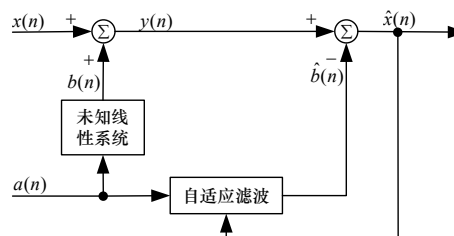


图 1 经典自适应滤波降噪

在机器人上的噪声消除应用中, 由于麦克风阵列受体积限制, 每个阵元采集的信号都包含原始信号和噪声信号, 从而不能将某个阵元采集的信号作为参考噪声, 因此无法进行有效的自适应滤波。文献[5]提出广义旁瓣抵消方法, 尝试利用空间信息形成波束和自适应方法相结合来解决上述问题, 但其实际性能受到时延估计和阻塞矩阵的影响<sup>[6]</sup>, 实际应用场合非常有限。

本文针对此问题, 利用人形机器人具有的类似人体的头相关传输函数的性质, 结合自适应滤波方法, 对基于空间信息的方法进行扩展: 对一路或多路阵元采集的信号按信道间差异特性进行处理, 降低其与目标声源方向原始声音信号的相关性, 得到与目标信号弱相关的参考噪声, 然后进行自适应滤波, 得到对目标信号的实时估计。

**作者简介:** 林剑峰(1978-), 男, 硕士研究生, 主研方向: 数字音频处理; 冯 瑞, 副教授、博士

**收稿日期:** 2008-07-20 **E-mail:** 052021197@fudan.edu.cn

## 2 信道间差异特性的噪声消除方法

麦克风阵列挂载在人形机器人身上, 考虑到躯干散射和阵元挂载位置的影响, 阵元采集信号时具有类似于人的双耳听觉立体性的特性, 在阵元间存在稳定的差异系统<sup>[7]</sup>, 因此, 可以利用差异特性进行噪声消除, 实现目标声音的增强。

### 2.1 问题描述

如图 2 所示, 以机器人为原点建立极坐标系,  $r$  表示距离,  $\theta$  表示水平方位角, 机器人正面为  $0^\circ$ ,  $\varphi$  表示仰角, 水平面为  $0^\circ$ 。

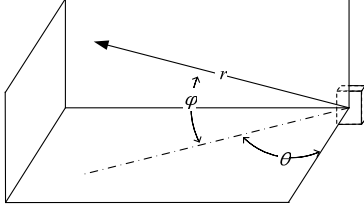


图 2 以机器人为中心的极坐标系

机器人上装有  $M$  个麦克风组成的阵列, 各阵元采集到的信号分别记为  $y_j$ ,  $j \in [1, M]$ 。设位于  $K$  个处于不同位置的声源  $s_i$ ,  $i \in [1, K]$  发出均值为零、互不相关(或弱相关)的声音信号, 第  $j$  个麦克风采集的信号  $y_j$  中由  $s_i$  产生的分量为  $x_{i,j}$ , 则有:

$$y_j(n) = \sum_{i=1}^K x_{i,j}(n) \quad (1)$$

针对声源  $s_1$  的声音增强问题等价于从  $y_j$  中将来自于其他方向声源的分量消除, 即: 已知  $y_j(n)$  和  $M, K$  未知, 求  $x_{1,j}(n)$ ,  $j \in [1, M], n \in [0, N-1]$ 。

### 2.2 头相关传输函数和信道间差异系统

头相关传输函数(Head Related Transfer Function, HRTF)描述了声源和人耳之间的传输信道的特征<sup>[8]</sup>。由于头部、躯干、耳廓对声波的散射以及声波在传输过程中的时延和衰减, 人耳听到的声音并不直接等同于来自声源的原始信号。对于人形机器人, 从声源到机器人的麦克风阵列阵元的声传播信道也具有类似特征。

参考人体的 HRTF 定义<sup>[9]</sup>, 对于给定的机器人, 头相关传输函数定义为

$$H_j(r, \theta, \varphi, \omega) = P_j(r, \theta, \varphi, \omega) / P_0(r, \omega), \quad j \in [1, M] \quad (2)$$

其中,  $\omega$  是声波频率;  $P_j$  是麦克风阵列第  $j$  个阵元处的声压;  $P_0$  是机器人不存在时原点处声压。由式(2)可以看出, HRTF 是与声源位置和声波频率相关的函数, 进一步由于远场效应, 当  $r$  较大时(1.2 m 以上)可以忽略其影响, 因此在机器人应用中, 可忽略  $r$  的影响。

根据传输函数的定义, 可以推出下式:

$$X_{i,j}(\omega) = S_i(\omega) H_j(\theta_i, \varphi_i, \omega) \quad (3)$$

其中,  $X_{i,j}(\omega)$ ,  $S_i(\omega)$  分别是  $x_{i,j}(n)$  和  $s_i(n)$  的频域表达, 记:

$$D_{i,p,q} = \frac{X_{i,p}(\omega)}{X_{i,q}(\omega)} = \frac{H_p(\theta_i, \varphi_i, \omega)}{H_q(\theta_i, \varphi_i, \omega)} \quad (4)$$

则  $D_{i,p,q}$  描述了声源  $i$  到阵元  $p$  和阵元  $q$  信道之间的差异, 称为基于 HRTF 的信道间差异传输函数。根据式(1)和式(4), 可以得到:

$$\begin{pmatrix} X_{1,1} & X_{2,1} & \cdots & X_{K,1} \end{pmatrix} \begin{pmatrix} 1 & D_{1,2,1} & \cdots & D_{1,M,1} \\ 1 & D_{2,2,1} & \cdots & D_{2,M,1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & D_{K,2,1} & \cdots & D_{K,M,1} \end{pmatrix} = (Y_1 \ Y_2 \ \cdots \ Y_M) \quad (5)$$

其中,  $Y_j$  是  $y_j(n)$  的频域形式。

## 2.3 差异系统噪声消除的计算方法

虽然根据式(5)在  $K < M$  的条件下, 可以通过求解方程组分解出目标声源的信号, 但是因为  $K$  未知, 且实时求解线性方程组代价巨大。因此, 采用先分解出与信号无关的噪声参考, 然后自适应对消的方法。由于声波散射时受波长与障碍物尺寸之间关系的影响, 因此在面对广谱声音时, 很难准确地对信道间差异系统传输函数  $D$  进行实时估计。并且由于  $D$  是非时变的, 主要依赖于机器人本身, 因此可以采取预先训练的策略: 将空间划分为  $L$  个方向, 在静音环境中, 依次在每个方向设置声源, 按下式求出一组值:

$$d_{\theta,\varphi,p,q}(\omega) = \frac{X_{1,p}(\omega)}{X_{1,q}(\omega)} = \frac{Y_p(\omega)}{Y_q(\omega)} \quad (6)$$

其中,  $d_{\theta,\varphi,p,q}(\omega)$  代表声源方向为  $(\theta, \varphi)$  时阵元  $p$  和阵元  $q$  之间的差异函数;  $Y_p(\omega)$  和  $Y_q(\omega)$  分别是阵元  $p$  和  $q$  采集到的信号的离散傅里叶变换;  $X_{1,p}(\omega)$  表示来自方向为  $(\theta, \varphi)$  唯一声源在阵元  $p$  中的分量的离散傅里叶变换。

训练完成后, 在运行时计算框架如图 3 所示。采集到的多路信号输入到系统中, 选择其中一路作为主信号源, 在这里以第 1 路  $y_1(n)$  为主信号源。其他采集到的信号和目标声源的方向输入滤波器  $D$ , 产生参考噪声  $r(n)$ , 然后经自适应滤波得到噪声消除后的信号。

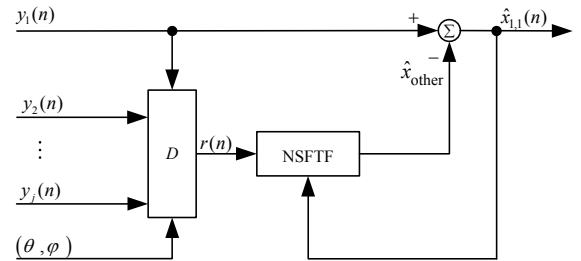


图 3 基于差异传输函数的自适应降噪

对于  $D$  系统, 根据输入的声源方向  $(\theta_0, \varphi_0)$ , 选择对应的  $d_{\theta,\varphi,1,q}(\omega)$ , 记  $d_{1,q} = d_{\theta,\varphi,1,q}(\omega)|_{\theta=\theta_0, \varphi=\varphi_0}$ ,  $Y_i = FFT[y_i(n)]$ , 则定义  $D$  系统为

$$r(n) = FFT^{-1} \left[ \sum_{i=2}^M (Y_i - Y_1 \times d_{1,i}) \right] \quad (7)$$

根据声源之间互不相关的特性以及式(1)和式(4), 可以得到:

$$r(n) = FFT^{-1} \left[ \sum_{i=2}^M \sum_{k=2}^K X_{k,i} (1 - d_{1,i}) \right] = \sum_{i=2}^K u_i x_{i,1} \quad (8)$$

其中,  $u_i$  为未知系数。因此, 当  $u_i \neq 0$  时,  $r(n)$  与  $\sum_{i=2}^K x_{i,1}$  相关, 与  $x_{1,1}$  无关, 则可根据最小二乘方准则得到通道 1 中的噪声估计  $\hat{x}_{\text{other}} = \sum_{i=2}^K \hat{x}_{i,1}(n)$ , 经消减后得到目标信号在信道 1 中分量的估计  $\hat{x}_{1,1}(n)$ 。自适应滤波采用数值稳定的快速递归最小二乘方算法 NSFTF(Numerical Stable Fast Transversal Filter)<sup>[10]</sup>。

## 3 仿真实验

### 3.1 信号信噪比改善性能实验

仿真试验中模仿了 2 个方位角夹角为  $60^\circ$  的声源, 与机器人的麦克风阵列距离为 2.5 m, 仰角均为  $0^\circ$ 。2 个声源产生不同的声音, 其中, 一个声源作为信号源, 另一个作为干扰噪声源。表 1 给出了本方法在不同噪声干扰级别上的信噪比提升结果以及与基于递归最小二乘法(Recursive Least Square, RLS)的传统噪声对消方法进行了比较。

表1 信噪比提升比较 dB

原信噪比	本方法信噪比提升	RLS方法信噪比提升
10	6	3
0	12	4
-10	15	3

### 3.2 语音识别性能改善实验

实验中语音识别采用的是为娃娃机器人研制的独立词识别系统。使用多个扬声器模拟具有混响特性的会场环境的背景噪音，逐次提升音量，在各个不同的信噪比级别上作对比测试。测试使用录制的100个口语命令，在机器人正面2.5m处播放，以(正确识别的命令数/100)作为识别正确率。实验结果如表2所示。

表2 语音识别性能提升

信噪比/dB	直接识别率/(%)	降噪后识别率/(%)
25	100	100
10	91	95
0	67	91
-10	31	79

## 4 结束语

本文提出的用于人形机器人的定向声音信号增强方法，有效地利用了人形机器人的麦克风阵列各阵元采集信号的传输差异性，能够实时地提升特定方向声音的信噪比，改善语音识别的效果，使机器人能够在有噪声干扰的条件下正确识别使用者的口语命令。

### 参考文献

[1] Nakadai K, Okuno H G, Kitano H. Robot Recognizes Three Simultaneous Speech by Active Audition[C]//Proc. of IEEE

International Conference on Robotics and Automation. [S. l.]: IEEE Press, 2003: 14-19.

[2] Valin J M, Rouat J, Michaud F. Enhanced Robot Audition Based on Microphone Array Source Separation with Post-filter[C]//Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems. [S. l.]: IEEE Press, 2004: 2123-2128.  
 [3] 许肖梅. 声学基础[M]. 北京: 科学出版社, 2003.  
 [4] 方艳梅, 刘永清. 数字信号处理[M]. 4版. 北京: 电子工业出版社, 2007.  
 [5] Gannot S, Cohen I. Speech Enhancement Based on the General Transfer Function GSC and Postfiltering[J]. IEEE Transactions on Speech and Audio Processing, 2004, 12(6): 561-571.  
 [6] 马晓红. 传声器阵列语音增强中关键技术的研究[D]. 大连: 大连理工大学, 2006.  
 [7] Algazi V R, Duda R O, Duraiswami R, et al. Approximating the Head-related Transfer Function Using Simple Geometric Models of the Head and Torso[J]. J. Acoust. Soc. Am., 2002, (112): 2053-2064.  
 [8] Gardner E G, Martin K D. HRTF Measurements of a KEMAR[J]. J. Acoust. Soc. Am., 1995, (99): 3907-3915.  
 [9] 钟小丽, 谢菠菘. 头相关传输函数的研究进展(1)[J]. 电声技术, 2004, (12): 44-46.  
 [10] Cioffi J M, Kailath T. Recursive-least-squares Transversal Filters for Adaptive Filtering[J]. IEEE Transactions on Acoustics, Speech, and Signal Processing, 1984, 32(2): 304-337.

编辑 索书志

(上接第176页)

在Step3中,  $U$ 实现对 $S$ 的身份验证, 并且实现了会话密钥的协商, 达到了一次一密钥的目的,  $U$ 对 $S$ 的身份认证建立在随机数选择的机密性和对称密钥算法的安全性上。此步骤中 $U$ 产生的一次性口令 $H^{N-i}(PW // Se)$ 用协商好的会话密钥 $K_i$ 加密传送, 提高了口令传送的安全性。

在Step4中,  $S$ 通过计算得到会话密钥 $K_i$ , 用 $K_i$ 解密收到的密文得到 $H^{N-i}(PW // Se)$ , 并对其鉴别, 此过程实现了 $S$ 对 $U$ 的身份认证。并且在认证通过后的正常通信中都用 $K_i$ 加密信息进行通信。

新型S/KEY方案的每次认证口令和会话密钥都不同, 并且一次有效, 因此, 对于重放攻击有很强的抵御能力。新方案实现了对 $S$ 的有效认证, 可以抵抗冒充攻击, 这里假定 $U$ 与攻击者控制的假冒服务器进行认证流程, 在Step1中,  $U$ 传送 $ID, K_{US}(Y_U)$ 给假冒服务器, 因为假冒服务器不知道秘密密钥 $K_{US}$ , 无法得到 $Y_U$ , 从而不能成功伪造生成 $R, R_1, R_2, K_{US}(Y_S)$ , 所以认证不能进行。

新型S/KEY方案可以有效抵御小数攻击, 因为种子值 $Se$ 和序列数 $Seq$ 都是经过随机数 $Y_U, Y_S$ 异或运算加密后传送, 攻击者无法得到 $Y_U, Y_S$ 的值, 不能篡改或伪造 $Se$ 和 $Seq$ , 所以不能进行小数攻击。

协议破坏攻击与中间人攻击在攻击S/KEY系统时有一个共同点, 就是一直窃听 $U$ 与 $S$ 的通信, 直到在关键步骤Step3中得到有效口令的密文, 然后利用此有效口令密文冒充合法用户继续与服务器进行认证, 虽然有效口令密文攻击者无法解密, 但是这里攻击者无须解密出明文的有效口令即可通过认证, 在新方案中攻击者无法破解获得会话密钥 $K_i$ , 在

Step4以后的通信过程通信双方的信息都必须用会话密钥加密传输, 攻击者无法得到会话密钥就不能进行通信, 因此, 新方案对这2种攻击方式有很强的抵抗能力。

## 6 结束语

新型S/KEY认证方案实现了共享密钥和会话密钥的协商, 会话密钥每次认证都重新协商产生, 实现了客户端与服务器之间双向认证, 确保了通信双方身份的真实有效。改进方案可以适合在安全性要求高的环境中使用, 尤其是在完善的认证授权中心比较少的情况下, 使用这种无需第三方认证机构、认证流程简单的新型S/KEY认证方案更加切实有效, 因此, 该认证方案具有一定的推广价值。

### 参考文献

[1] Lamport L. Password Authentication with Insecure Communication[J]. Communications of the ACM, 1981, 24(11): 770-772.  
 [2] Haller N M. The S/Key One-time Password System[C]//Proceedings of the Internet Society Symposium on Network and Distributed System Security. San Diego, CA, USA: [s. n.], 1994.  
 [3] 高雪, 张焕国, 孙晓梅. 一种改进的一次性口令认证方案[J]. 计算机应用研究, 2006, 23(6): 127-128.  
 [4] 梁喆, 方勇. 融入混沌理论的一种OTP方案研究[J]. 计算机工程, 2007, 33(7): 160-162.  
 [5] Schneier B. Applied Cryptography-protocols Algorithms and Source Code in C[M]. New York, USA: John Wiley and Sons Inc., 1996.

编辑 顾姣健

